JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 109, 311–324 (1985)

Monotone Value Iteration for Discounted Finite Markov Decision Processes

D. J. WHITE

Department of Decision Theory, University of Manchester, Faculty of Economic and Social Studies, Manchester, MI3 3PL, England

Submitted by E. Stanley Lee

1. INTRODUCTION

In this paper we will consider value iteration for the following class of Markov decision process, viz., there is

(i) a finite set of states $I = \{i = 1, 2, ..., N\}$;

(ii) a finite set of actions $K(i)$ for each state $i \in I$;

(iii) an immediate reward r_i^k for $i \in I$, $k \in K(i)$;

(iv) a transition probability p_{ij}^k for each $i \in I$, $j \in I$, $k \in K(i)$;

(v) a discount factor ρ, $0 \leqslant \rho < 1$.

The objective is to maximise the infinite horizon expected discounted reward beginning in any state.

It is well known (see, e.g., [1]) that the maximal expected discounted rewards, $\{v(i)\}$, beginning in any state $i \in I$, are a unique solution to the equation

$$v(i) = \max_{k \in K(i)} \left[r_i^k + \rho \sum_{j \in I} p_{ij}^k v(j) \right], \qquad \forall i \in I \tag{1}$$

and that if $\delta = (k(1), k(2), ..., k(N))$ are optimal values of k in (1), then the repeated use of the decision rule δ will give the maximal expected discounted return for each state $i \in I$.

The standard value iteration method for solving (1) (or at least approximating the solution to (1)) is as follows (e.g., see [1]).

$$v_0 = u$$

$$n \geqslant 1: \quad v_n = Tv_{n-1}$$

0022-247X/85 $3.00

where, for any $z: I \to \mathscr{R}^1$,

$$[Tz]_i = \max_{k \in K(i)} [T^k z]_i$$

$$[T^k z]_i = r_i^k + \rho \sum_{j \in I} p_{ij}^k z(j).$$

The properties of this method are well known (e.g., see [1]).

Although this method has been thoroughly studied, it is by no means the only method of solving (1). In [2], for example, a simplicial method for calculating fixed points is suggested. In [3], a Gauss–Seidel approach is used, which is a modification of the standard value iteration. In [4], a survey is made of various methods.

There is no a priori reason why the standard value iteration method should even be a good method. One can envisage a general class of methods of the form

$$v_n = \theta v_{n-1}$$

where θ is by no means restricted to T.

The purpose of this paper is to examine two variants of T, which seem to have some intuitive value, but whose final assessment must reside in proper numerical analysis and testing. It is not the purpose of this paper to do the latter, rather this is merely an "ideas" paper.

The two methods which we will study are as follows:

Case 1. $\theta = \theta_1$

$$v_0 = u: I \to \mathscr{R}^1.$$

$n \geqslant 1$: $\theta_1 v_{n-1} = T_1(n)\, v_{n-1}$

$$[T_1(n)\, z]_i = \max_{k \in K(i)} [T_1^k(n)\, z]_i, \qquad z: I \to \mathscr{R}^1$$

$$[T_1^k(n)\, z]_i = r_i^k + \rho(n) \sum_{j \in I} p_{ij}^k z(j)$$

$$\rho(n) = \lambda(n)\, \rho$$

$$\lambda(n) \geqslant 0.$$

Case 2. $\theta = \theta_2$

$$v_0 = u: I \to \mathscr{R}^1.$$

$n \geqslant 1$: $\theta_2 v_{n-1} = T_2(n)\, v_{n-1}$

$$[T_2(n)\, z]_i = \max_{k \in K(i)} [T_2^k(n)\, z]_i, \qquad z: I \to \mathscr{R}^1$$

$$[T_2^k(n)\, z]_i = r_i^k + \rho \sum_{j \in I} p_{ij}^k z + \varepsilon(n)$$

$$\varepsilon(n) \geqslant 0.$$

In [5] a study of the convergence properties of such iterations is given. In [5] the motive is not really a computational one for solving (1), and the replacement of ρ by $\rho(n)$ and r_i^k by $r_i^k + \varepsilon(n)$ is to be interpreted as a move from stationary problems to time-dependent problems. The convergence analysis will naturally apply to our iterations of course. However, we will be interested in conditions for which the convergence will be monotone and in features not covered in [5].

In addition, it is clearly possible to generalise the method by making $\{\lambda(n)\}$ and $\{\varepsilon(n)\}$ dependent on i. The paper is exploratory and, for this reason, it is kept as simple as possible.

Finally, we will assume that $r_i^k \geqslant 0$ $\forall i \in I$, $k \in K(i)$, and that $u \geqslant 0$. The first restriction may be made without loss of generality (see [1]), and both restrictions ensure that $v_n \geqslant 0$, $\forall n$.

2. Theory

2.1. *Case* θ_1

We will study the problem under the following conditions.

C_1.

$$n \geqslant 2: \quad \lambda(n) = \max[1, \lambda(n-1)\, \tau(n-1)]. \tag{2}$$

$$n = 1: \quad \lambda(1) \geqslant 1 \tag{3}$$

$$T_1(1)\, u \geqslant u \tag{4}$$

$$p\lambda(1)\, \tau(1) \geqslant 1 \tag{5}$$

where

$$\tau(n-1) = \max_{i \in I} [v_{n-2}(i)/v_{n-1}(i)]$$

with $v_{n-2}(i)/v_{n-1}(i) = 1$ if $v_{n-2}(i) = v_{n-1}(i) = 0$.

We then have the following theorem.

Theorem 1. *Under the conditions* C_1 *the following results hold.*

(i) *The sequence* $\{v_n\}$ *will form an increasing sequence*;

(ii) *the sequence* $\{\lambda(n)\}$ *will form a decreasing sequence converging to some* $\lambda \geqslant 1$;

(iii) *the sequence* $\{v_n\}$ *will converge to a unique solution* v *of Eq.* (1) *with* ρ *replaced by* $\lambda\rho$, *i.e.*,

$$v(i) = \max_{k \in K(i)} \left[r_i^k + \rho\lambda \sum_{j \in I} p_{ij}^k v(j) \right]; \tag{6}$$

(iv) *if, for any specified* u, $\{\tilde{v}_n\}$, $\{v_n\}$ *are the sequences given by the* T, θ_1 *iterative methods, respectively, then*

$$v_n \geqslant \tilde{v}_n, \qquad \forall n \tag{7}$$

$$v \geqslant \tilde{v} \tag{8}$$

where $\tilde{v}$ *is the solution to* (1);

(v) *if*

$$l = \min_{i \in I}[u(i)], \qquad m = \min_{i \in I, k \in K(i)}[r_i^k]$$

$$L = \max_{i \in I}[u(i)], \qquad M = \max_{i \in I, k \in K(i)}[r_i^k]$$

$$a = \rho(\lambda(1) - 1)\, m/(1-\rho), \qquad b = \rho(\lambda(1) - 1)\, M/(1-\rho)$$

then, $\forall n \geqslant 1$, $i \in I$,

$$\begin{aligned} v_n(i) - b(1 + \rho\lambda(n) + \rho^2\lambda(n)(n-1) + \rho^{n-1}\lambda(n)\,\lambda(n-1)\cdots\lambda(2)) \\ - \rho^n\lambda(n)\,\lambda(n-1)\cdots\lambda(1)(L - m/(1-\rho)) \\ \leqslant v(i) \leqslant v_n(i) - a(1 + \rho\lambda(n) + \rho^2\lambda(n)\,\lambda(n-1) \\ + \cdots \rho^{n-1}\lambda(n)\,\lambda(n-1)\cdots\lambda(2)) \\ + \rho^n\lambda(n)\,\lambda(n-1)\cdots\lambda(1)(M/(1-\rho) - l) \end{aligned} \tag{9}$$

where

$$\rho^t\lambda(n)\,\lambda(n-1)\cdots\lambda(n-t+1) \leqslant (\rho\lambda(1)\,\tau(1))^t$$

and hence the left-hand and right-hand sides of the inequalities will converge;

(vi) *if* $T_1(1)\,u \geqslant \lambda(1)\,u$, *then*

$$\lambda(2) = 1, \qquad \lambda = 1, \qquad \tilde{v} = v. \tag{10}$$

Note that when $u \geqslant 0$, as in the case in this paper, condition (10) *will imply condition* (4).

Proof. (i) For $n \geqslant 2$, $i \in I$,

$$v_n(i) - v_{n-1}(i) = [T_1(n)\, v_{n-1}]_i - [T_1(n-1)\, v_{n-2}]_i$$

$$\geqslant \min_{k \in K(i)} \left[\rho \sum_{j \in I} p_{ij}^k (\lambda(n)\, v_{n-1}(j) - \lambda(n-1)\, v_{n-2}(j)) \right] \geqslant 0,$$

from (2), providing $v_{n-1} \geqslant v_{n-2}$.

Hence the requisite result follows, providing $v_1 \geqslant v_0 = u$, and this follows from (4).

(ii) From (i), for $n \geqslant 2$, we have $\tau(n-1) \leqslant 1$, and hence from (2) we have $\lambda(n) = 1$ or $\lambda(n) \leqslant \lambda(n-1)$. The requisite result now follows since any decreasing bounded sequence converges and each $\lambda(n) \geqslant 1$.

(iii) Let

$$M = \max_{i \in I, k \in K(i)} [r_i^k]$$

$$L = \max_{i \in I} [u(i)].$$

Then, for $i \in I$, $n \geqslant 1$,

$$v_n(i) \leqslant M \sum_{t=0}^{n-1} \rho^t \prod_{S=0}^{t} \lambda(t) + L\rho^n \prod_{s=0}^{n} \lambda(s).$$

Now, for $t \geqslant 2$,

$$\begin{aligned} \lambda(t) &= \max \left[1, \prod_{s=1}^{t-1} \tau(s) \cdot \lambda(1) \right] \\ &\leqslant \max[1, \tau(1)\, \lambda(1)]. \end{aligned} \tag{11}$$

Hence $\{v_n\}$ will be bounded if $\rho\lambda(1)\, \tau(1) < 1$, which is specified in (5).

$\{v_n\}$ is monotone increasing and is bounded, and hence converges to some v. Since $\{\lambda(n)\}$ also converges, it is now a trivial problem to show that v satisfies the requisite equation. Note that $\lambda\rho < 1$.

(iv) For $n \geqslant 1$, $i \in I$,

$$v_n(i) - \tilde{v}_n(i) = [T_1(n)\, v_{n-1}]_i - [T\tilde{v}_{n-1}]_i$$

$$\geqslant \min_{k \in K(i)} \left[(\lambda(n) - 1)\, \rho \sum_{j \in I} p_{ij}^k v_{n-1}(j) \right.$$

$$\left. + \rho \sum_{j \in I} p_{ij}^k (v_{n-1}(j) - \tilde{v}_{n-1}(j)) \right].$$

Since $v_0 - \tilde{v}_0 = u - u = 0$ and $\lambda(n) \geqslant 1$, an inductive analysis will give the requisite result.

The fact that $v \geqslant \tilde{v}$ is obvious.

(v) For $n \geqslant 1$, $i \in I$,

$$\begin{aligned} v_n(i) - \tilde{v}(i) &= [T_1(n)\, v_{n-1}]_i - [T\tilde{v}]_i \\ &\geqslant \min_{k \in K(i)} \left[\rho\lambda(n) \sum_{j \in I} p_{ij}^k v_{n-1}(j) - \rho \sum_{j \in I} p_{ij}^k \tilde{v}(j) \right] \\ &\geqslant \min_{k \in K(i)} \left[\rho\lambda(n) \sum_{j \in I} p_{ij}^k (v_{n-1}(j) - \tilde{v}(j)) \right] \\ &\quad + \min_{k \in K(i)} \left[\rho(\lambda(n) - 1) \sum_{j \in I} p_{ij}^k \tilde{v}(j) \right]. \end{aligned}$$

Let

$$\nabla_n = \min_{i \in I} [v_n(i) - \tilde{v}(i)].$$

Then

$$\begin{aligned} \nabla_n &\geqslant \rho\lambda(n)\, \nabla_{n-1} + \rho(\lambda(n) - 1)\, m/(1 - \rho) \\ &\geqslant \rho\lambda(n)\, \nabla_{n-1} + a \end{aligned}$$

where

$$a = \rho(\lambda(1) - 1)\, m/(1 - \rho).$$

Then

$$\begin{aligned} \nabla_n &\geqslant a(1 + \rho\lambda(n) + \rho^2\lambda(n)\, \lambda(n-1) + \cdots \rho^{n-1}\lambda(n)\, \lambda(n-1) \cdots \lambda(2)) \\ &\quad + \rho^n\lambda(n)\, \lambda(n-1) \cdots \lambda(1)\, \nabla_0 \\ &\geqslant a(1 + \rho\lambda(n) + \rho^2\lambda(n)\, \lambda(n-1) + \cdots \rho^{n-1}\lambda(n)\, \lambda(n-1) \cdots \lambda(2)) \\ &\quad + \rho^n\lambda(n)\, \lambda(n-1) \cdots \lambda(1)(l - M/(1 - \rho)). \end{aligned}$$

Similarly if

$$\Delta_n = \max_{i \in I} [v_n(i) - \tilde{v}(i)]$$

we have

$$\begin{aligned} \Delta_n &\leqslant b(1 + \rho\lambda(n) + \rho^2\lambda(n)\, \lambda(n-1) + \cdots \rho^{n-1}\lambda(n)\, \lambda(n-1) \cdots \lambda(2)) \\ &\quad + \rho^n\lambda(n)\, \lambda(n-1) \cdots \lambda(1)\, \lambda(L - m/1 - \rho)) \end{aligned}$$

where

$$b = \rho(\lambda(1) - 1)\, M/(1-\rho).$$

The requisite result now follows, noting that the convergence result follows from (11), (5).

(vi) $$\lambda(2) = \max[1, \lambda(1)\, \tau(1)].$$

Now

$$\tau(1) = \max_{i \in I} \left[u(i) \Big/ \max_{k \in K(i)} \left[r_i^k + \rho\lambda(1) \sum_{j \in I} p_{ij}^k u(j) \right] \right].$$

From (9) we have $\tau(1) \leqslant 1/\lambda(1)$. Hence $\lambda(2) = 1$, and the rest follows. ∎

We make the following observations.

1. In (v), the standard value iteration is equivalent to $\lambda(1) = 1$, and hence $a = 0$, $b = 0$, and the inequality reduces to the standard result (see [1] with $m = 0$). It is not clear exactly how the levels compare with the standard case and these remain to be studied.

2. A lot would depend on the convergence properties of the sequence $\{\lambda(n)\}$. In (vi) we have $\lambda(2) = \lambda = 1$. In this case (satisfied always when $u = 0$), the method is always at least as good as the standard method, at least in terms of the iterations, making use of (iv). For a general u there is the extra calculation required to ensure that (10) holds, and there is the extra work involved in calculating $\{\tau(n)\}$ and $\{\lambda(n)\}$.

If more general conditions could be determined for which $\{\lambda(n)\}$ converges to $\lambda = 1$, then again, purely from the iterative point of view, the modified θ_1 method will give better approximations if $\{v_n\}$ are used as estimates.

Expression (11) gives $\lambda(n)$ in terms of $\{\tau(t)\}_1^{n-1}$. We always have $\tau(t) \leqslant 1$, and it seems quite possible that for many cases $\{\lambda(n)\}$ will converge to λ fairly close to 1.

This is not universally true. Thus, suppose $\rho\lambda(1) < 1$, $\lambda(1) > 1$, and u is the unique solution to

$$u = T_1(1)\, u.$$

Then

$$v_1 = T_1(1)\, u = u$$

$$\tau(1) = 1, \qquad \lambda(2) = \lambda(1)$$

and it is easily seen that $v_n = u$, $\tau(n) = 1$, $\lambda(n) = \lambda(1)$, $\forall n \geqslant 1$, and hence $\{\lambda(n)\}$ converges to $\lambda = \lambda(1) > 1$.

3. As a special case of u, we could take $u(i) = l$, $\forall i \in I$. From (4) we require that if, for comparison purposes, we wish this to apply for the standard case, if $q = \min_{i \in I} \max_{k \in K(i)} [r_i^k]$ then

$$l \leqslant q/(1-\rho).$$

If we then select $\lambda(1)$ to satisfy

$$1 \leqslant \lambda(1) \leqslant q/l(1-\rho)$$

condition (10) will hold and the modified θ_1 method will be at least as good as the standard method for such u.

Finally, to complete this section we use a simple problem to illustrate the method.

We take $\rho = 0.9$, $\lambda(1) = 1.1$, and the remaining data as follows.

It will be noted that we have allowed negative $\{r_i^k\}$. This is quite permissible, as the proofs will show, providing $u \geqslant 0$, since it is really the non-negativity of $\{v_n\}$ we wish to preserve. The proofs are, of course, easily adapted for negative $\{r_i^k\}$, for example, by first of all transforming to non-negative $\{r_i^k\}$.

i	k	r_i^k	p_{i1}^k	p_{i2}^k
1	1	6	0.5	0.5
	2	4	0.8	0.2
2	1	−3	0.4	0.6
	2	−5	0.7	0.3

$$u(1) = 15.5, \qquad u(2) = 5.60.$$

This corresponds to the policy $k(1) = k(2) = 1$, and ensures that the standard method will give a monotone increasing sequence as well as the modified θ_1 method (i.e., condition (4) is satisfied for $\lambda(1) = 1.1$ and $\lambda(1) = 1$). Since $\rho\lambda(1) < 1$, condition (5) is also satisfied. It is to be noted that condition (10) is not satisfied (see tabulations below), although the results of (vi) are satisfied. The tabulations are as follows.

θ_1 method.

$n = 1$: $\lambda(1) = 1.1$, $v_1(1) = 17.4$, $v_1(2) = 7.4$, $\tau(1) = 0.88$, $\lambda(1)\,\tau(1) = 0.97$.

$n = 2$: $\lambda(2) = 1$. Hence the method now continues as with the standard method.

Standard method.

$n = 1$: $\lambda(1) = 1$, $\tilde{v}_1(1) = 16.2$, $\tilde{v}_1(2) = 6.3$.

Optimal solution.

$$\tilde{v}(1) = 22.2,\ \tilde{v}(2) = 11.9,\ k(1) = 2,\ k(2) = 2.$$

We have not completed all the calculations since it is enough to show that in this case the modified θ_1 method is clearly superior to the standard method for the chosen circumstances.

2.2. *Case* θ_2

We will study the problem under the following conditions.

C_2.

$$n \geqslant 2: \quad \varepsilon(n) = \max[\varepsilon(n-1) - \rho\eta(n-1), 0]. \tag{12}$$

$$n = 1: \quad \varepsilon(1) \geqslant 0 \tag{13}$$

$$T_2(1)\, u \geqslant u \tag{14}$$

where

$$\eta(n-1) = \min_{i \in I}[v_{n-1}(i) - v_{n-2}(i)].$$

We then have the following theorem.

THEOREM 2. *Under conditions* C_2 *the following results hold.*

(i) *The sequence* $\{v_n\}$ *will form an increasing sequence;*

(ii) *the sequence* $\{\varepsilon(n)\}$ *will form a decreasing sequence converging to some* $\varepsilon \geqslant 0$;

(iii) *the sequence* $\{v_n\}$ *will converge to a unique solution of Eq.* (1) *with the addition of the term* ε, *viz.*,

$$v(i) = \max_{k \in K(i)} \left[r_i^k + \varepsilon + \rho \sum_{j \in I} p_{ij}^k v(j) \right]; \tag{15}$$

(iv) *if, for any specified u,* $\{\tilde{v}_n\}$, $\{v_n\}$ *are the sequences given by the T,* θ_2 *iterative methods respectively, then*

$$v_n \geqslant \tilde{v}_n, \qquad \forall n \tag{16}$$

$$v \geqslant \tilde{v} \tag{17}$$

where $\tilde{v}$ *is the solution to* (1);

(v)

$$v_n(i) - \sum_{t=0}^{n-1} \rho^t \varepsilon(n-t) - \rho^n(L - m/(1-\rho))$$

$$\leqslant \tilde{v}(i) \leqslant v_n(i) - \sum_{t=0}^{n-1} \rho^t \varepsilon(n-t) + \rho^n(M/(1-\rho) - l);$$

(iv) *if*

$$T_2(1)\, u \geqslant u + \varepsilon(1)\, e/\rho \tag{18}$$

where $e \in \mathscr{R}^m$ *has unit components, then* $\varepsilon(2) = 0$, $\varepsilon = 0$, $\tilde{v} = v$.

Note that (18) *implies* (14).

Proof. (i) For $n \geqslant 2$, $i \in I$,

$$v_n(i) - v_{n-1}(i) = [T_2(n)\, v_{n-1}]_i - [T_2(n-1)\, v_{n-2}]_i$$

$$\geqslant \min_{k \in K(i)} \left[\rho \sum_{j \in I} p_{ij}^k (v_{n-1}(j) - v_{n-2}(j)) \right] + \varepsilon(n) - \varepsilon(n-1)$$

$$\geqslant 0$$

from (12).

For $n = 1$, $i \in I$,

$$v_1(i) - v_0(i) = [T_2(1)\, u]_i - u(i) \geqslant 0$$

from (14), and the requisite result holds.

(ii) From (i) and (12) we have $\varepsilon(n) = 0$ or $\varepsilon(n) \leqslant \varepsilon(n-1)$. The requisite result now follows since any decreasing bounded sequence converges and each $\varepsilon(n) \geqslant 0$.

(iii) For $i \geqslant I$, $n \geqslant 1$,

$$v_n(i) \leqslant M \sum_{t=0}^{n-1} \rho^t + \rho^n L + \sum_{t=0}^{n-1} \rho^t \varepsilon(n-t)$$

$$\leqslant (M + \varepsilon(1))(1 - \rho^n)/(1-\rho) + \rho^n L.$$

$\{v_n\}$ will be bounded and hence converges to some v. Since $\{\varepsilon(n)\}$ also converges it is now a trivial problem to show that v satisfies the requisition equation.

(iv) For $n \geqslant 1$, $i \in I$,

$$v_n(i) - \tilde{v}_n(i) = [T_2(n)\, v_{n-1}]_i - [T\tilde{v}_{n-1}]_i$$
$$\geqslant \min_{k \in K(i)} \left[\rho \sum_{j \in I} p_{ij}^k (v_{n-1}(j) - \tilde{v}_{n-1}(j)) \right] + \varepsilon(n).$$

Since $v_0 - \tilde{v}_0 = u - u = 0$, and $\varepsilon(n) \geqslant 0$, inductive analysis will give the requisite result.

The fact that $v \geqslant \tilde{v}$ is obvious.

(v) For $n \geqslant 1$, $i \in I$,

$$v_n(i) - \tilde{v}(i) = [T_2(n)\, v_{n-1}]_i - [T\tilde{v}]_i$$
$$\geqslant \min_{k \in K(i)} \left[\rho \sum_{j \in I} p_{ij}^k (v_{n-1}(j) - \tilde{v}(j)) \right] + \varepsilon(n).$$

Then:

$$\nabla_n \geqslant \rho \nabla_{n-1} + \varepsilon(n)$$
$$\geqslant \sum_{t=0}^{n-1} \rho^t \varepsilon(n-t) + \rho^n (l - M/(1-\rho)).$$

Similarly

$$\Delta_n \leqslant \sum_{t=0}^{n-1} \rho^t \varepsilon(n-t) + \rho^n (L - m/(1-\rho)).$$

The requisite result now follows.

(vi)

$$\varepsilon(2) = \max[\varepsilon(1) - \rho\eta(1), 0].$$

Hence $\varepsilon(2) = 0$ if

$$\varepsilon(1) \leqslant \rho(v_1(i) - u(i)), \qquad \forall i \in I$$

i.e., $T_2(1)\, u \geqslant u + \varepsilon(1)\, e/\rho$.

The requisite result now follows. ∎

From this theorem we may make certain observations.

1. In (v) the standard value iteration is equivalent to $\varepsilon(1) = 0$ and hence $\varepsilon(n-t) = 0$, $\forall n, t$, and the result reduces to the standard result (see [1] with $l = L = m = 0$). The use of the modified θ_2 method will give a

tighter upper bound and a weaker lower bound. The exact behaviour would need to be studied experimentally.

2. A lot would depend on the convergence properties of the sequence $\{\varepsilon(n)\}$. In (vi) we have $\varepsilon(2)=\varepsilon=0$. In this case (not always satisfied by $u=0$, in contrast with the θ_1 method) the method is always at least as good as the standard method, at least in terms of the iterations, making use of (iv). For a general u there is the extra calculation required to ensure that (18) is satisfied, and there is the extra work in calculating $\{\eta(n)\}$, $\{\varepsilon(n)\}$.

If more general conditions could be determined for which $\{\varepsilon(n)\}$ converges to $\varepsilon=0$, then again, purely from an iterative point of view, the modified θ_2 method will give better approximations if $\{v_n\}$ are to be used as the estimates. Expression (12) may be used to give, for $n\geqslant 2$,

$$\varepsilon(n)=\max\left[0, \varepsilon(1)-\rho\sum_{t=1}^{n-1}\eta(t)\right]. \tag{19}$$

We always have $\eta(t)\geqslant 0$, and it seems quite possible that for many cases $\{\varepsilon(n)\}$ will converge to ε close to 0.

This is not universally true. Thus suppose $\varepsilon(1)>0$ and u is the unique solution to

$$T_2(1)\,u=u.$$

Then

$$v_1=T_2(1)\,u=u$$

$$\eta(1)=0, \qquad \varepsilon(2)=\varepsilon(1)$$

and it is easily seen that $v_n=u$, $\eta(n)=0$, $\varepsilon(n)=\varepsilon(1)$, $\forall n\geqslant 1$, and hence $\{\varepsilon(n)\}$ converges to $\varepsilon=\varepsilon(1)>0$.

3. As a special case of u we would take $u(i)=l$, $\forall i\in I$. From (14) we require that if, for comparison purposes, we wish this to apply for the standard case, with $q=\min_{i\in I}\max_{k\in K(i)}[r_i^k]$, then

$$l\leqslant q/(1-\rho).$$

If we then select (1) to satisfy

$$0\leqslant\varepsilon(1)\leqslant\rho q/(1-\rho)-\rho l$$

providing the right-hand side is non-negative, then condition (18) will hold and the modified θ_2 method will be at least as good as the standard method for such u.

Finally, as in Section 2.1, we use the same problem to illustrate the calculations, again allowing negative $\{r_i^k\}$. We set $\varepsilon(1) = 2$. The calculations are as follows, again with u as given in Section 2.2 which satisfies (14) with $\varepsilon(1) = 2$ and $\varepsilon(1) = 0$.

θ_2 method.

$n = 1$: $\varepsilon(1) = 2$, $v_1(1) = 18.2$, $v_1(2) = 8.3$, $\eta(1) = 2.7$.

$n = 2$: $\varepsilon(2) = 0$. Hence the method now continues as with the standard method.

Again we have not completed all the calculations since it is enough to show that in this case the θ_2 method is clearly superior to the standard method for the chosen circumstances.

In this case, as distinct from the θ_1 case, it is to be noted that, for this example, condition (18) holds, in which case $\varepsilon(2)$ must be 0.

3. Summary and Comments

This paper is an exploratory paper with the aim of studying two modifications of the standard value iteration method which maintain monotone behaviour of the sequence $\{v_n\}$. The two modifications are obtained in effect by a perturbation of the discount factor or of the rewards. Under fairly weak conditions given in Theorems 1 and 2 the methods produce a dominating sequence $\{v_n\}$, in the relationship to the sequence $\{\tilde{v}_n\}$ produced by the standard method of value iteration, for any specific starting solution u. If the inserted perturbations $\{\lambda(n)\}$ or $\{\varepsilon(n)\}$, in the two methods considered, converge respectively to 1 and 0, then not only do we have the requisite dominance but the methods give a sequence $\{v_n\}$ converging to $\tilde{v}$ and are uniformly better than the standard method for all u specified by the theorem.

Special conditions are given for which the requisite convergence of the $\{\lambda(n)\}$ or $\{\varepsilon(n)\}$ arises in two steps. This is useful, but excludes some u functions for which the requisite convergence will still hold. In order to enlarge the set of u for which the requisite convergence takes place it will be necessary to study the behaviour of the sequences $\{\lambda(n)\}$ or $\{\varepsilon(n)\}$ or of the related sequences $\{\tau(n)\}$ or $\{\eta(n)\}$. Even if $\{\lambda(n)\}$ does not converge to $\lambda = 1$, or $\{\varepsilon(n)\}$ does not converge to $\varepsilon = 0$, one might expect that the methods would work well for some u functions outside the specified sets. The facts that $\tau(n) \leqslant 1$, $\eta(n) \geqslant 0$, for all $n \geqslant 1$, coupled with the expressions (11), (19), give cause for some hope.

In any event, even if the convergent values λ or ε are not respectively 1 or 0, we can still obtain error bounds for v as with the standard method which are tighter on the upper bound and weaker on the lower bound in the case of θ_2, but this remains to be studied further in the case of θ_1.

REFERENCES

1. D. J. WHITE, "Finite Dynamic Programming," Wiley, New York, 1978.
2. J. F. SHAPIRO, Brouwers fixed point theorem and finite state space Markovian decision theory, *J. Math. Anal. Appl.* **49** (1975), 710–712.
3. H. KUSHNER, "Introduction to Stochastic Control," Holt, Rinehart & Winston, New York, 1971.
4. D. J. WHITE, A survey of algorithms for some restricted classes of Markov decision problems, *Proc. Operations Res.* **8** (1979), 103–121.
5. A. FEDERGRUEN AND P. J. SCHWEITZER, Nonstationary Markov decision problems with converging parameters, *J. Optim. Theory Appl.* **34** (1981), 207–242.